

Listening comprehension of academic and everyday language in first language and second language students

ALEXANDRA MARX

Humboldt University of Berlin and University of Potsdam

BIRGIT HEPPT and SOFIE HENSCHL

Humboldt University of Berlin

Received: October 2, 2015 Accepted for publication: May 17, 2016

ADDRESS FOR CORRESPONDENCE

Alexandra Marx, Department of Education, University of Potsdam, Karl-Liebknecht-Straße 24–25, Potsdam 14476, Germany. E-mail: alexandra.marx@uni-potsdam.de

ABSTRACT

Listening comprehension in academic contexts is an important prerequisite for the acquisition of school-related skills. However, the concept of academic language is generally the subject of much debate, and there is no consensus yet on what constitutes academic listening comprehension. To date, it remains unclear whether listening comprehension of academic language may be regarded as a distinct construct. The present study investigates the distinction of academic and everyday listening comprehension in 459 ninth-grade students speaking German as a first language or as a second language. Empirical results provide preliminary evidence that academic listening forms a separate facet of listening comprehension. Contrary to theoretical assumptions, second language students were not particularly disadvantaged in terms of academic listening comprehension.

Learning in any school subject depends on students' proficiency in listening comprehension in academic contexts (Rost, 1994). Classroom discourse has, for example, repeatedly been found to facilitate the acquisition of reading comprehension (cf. Nystrand, 2006; Wolf, Crosson, & Resnick, 2005). Consequently, listening comprehension in academic contexts may be regarded as a basis for the acquisition of knowledge and academic achievement.

For the most part, communication in academic contexts is marked by a high level of “writtenness,” whether information is presented written or orally. Academic listening texts possess characteristics of formal and planned discourse as they are lexically dense, context reduced, and characterized by textual structures that resemble those of written texts (Flowerdew & Miller, 1997). Classroom talk

may thus be localized at the interface of an oral–literate continuum where “oral” refers to prototypical speaking in everyday conversations and “literate” refers to prototypical writing in academic prose (Biber, 1995; Tannen, 1982). Although oral academic language is often less complex and contains fewer academic language features than written academic language (cf. Corson, 1997; Nagy & Townsend, 2012), oral and written academic language share important features, such as an informational focus and an abstract, impersonal style (Csomay, 2006). Oral language in academic settings is therefore more complex and more difficult to process than oral language in everyday conversations. Accordingly, several researchers have assumed that everyday listening and academic listening comprehension may rely on different sets of subskills and have proposed definitions of academic listening comprehension and its components (cf. Ferris & Tagg, 1996; Richards, 1983; Taylor & Geranpayeh, 2011). Most of these definitions are based on the presumption that academic listening comprehension requires the listener to concentrate on as well as understand long stretches of speech without being able to use the facilitating functions of interactive discourse (Flowerdew, 1994). For this particular type of listening comprehension, some authors propose specific comprehension subskills, such as the “differentiation between relevant and irrelevant information,” “command of strategies that support understanding (i.e., taking notes),” “the ability to identify the topic, purpose and scope of a lecture,” or “making inferences” (e.g., Buck, 2001; Flowerdew, 1995; Richards, 1983). Yet it remains unclear how these skills may be operationalized and to what extent they are specific for listening in academic contexts. To describe challenges of academic listening comprehension, it thus seems useful to resort to more general conceptualizations of academic language.

CHALLENGES OF ACADEMIC LANGUAGE

The general differentiation between language skills needed to interact with others in informal contexts and those needed to understand school-related language has been addressed by various researchers (e.g., Bailey & Butler, 2003; Schlepppegrell, 2004). An important conceptual basis of this line of research was the distinction between *basic interpersonal communication skills* (BICS) and *cognitive academic language proficiency* (CALP) proposed by Cummins (1980). According to Cummins (1980), BICS are defined as the language skills needed in social situations. When people interact with each other, their communication usually requires BICS because it occurs in a meaningful context and is thus relatively easy to understand and cognitively not very demanding. In contrast, CALP is needed to understand language in context-reduced settings and is cognitively demanding, such as when abstract ideas or concepts are presented. CALP thus primarily pertains to language typically used in academic contexts and can be defined as “the extent to which an individual has access to and command of the oral and written academic registers of schooling” (Cummins, 2000, p. 67).

The differentiation between BICS and CALP served as a basis for further conceptualizations of academic language. To date, there is no single definition of what constitutes academic language, and a variety of terms have been suggested to label the construct, such as “language of schooling,” “scientific language,” or

“advanced literacy” (cf. Snow & Uccelli, 2009). The US-based National Center for Research on Evaluation, Standards, and Student Testing (CRESST) has developed a conceptualization of academic language that adds additional features to the CALP definition of academic language (Bailey & Butler, 2003). According to this view, academic language can be defined as “the language that is used by teachers and students for the purpose of acquiring new knowledge and skills, imparting new information, describing abstract ideas, and developing students’ conceptual understanding” (Chamot & O’Malley, 1994, p. 40). On the content level, academic language is characterized by cognitively demanding school-related topics, such as lectures on abstract concepts or ideas (Bailey & Butler, 2003; Eckhardt, 2008). It is also context reduced to a certain extent: the information conveyed is often removed from its immediate context (i.e., there is a very limited amount of shared knowledge or context between speaker and listener; cf. Curen-ton & Justice, 2004).

Academic language is associated with the mastery of specific genres, command of specific reasoning and argumentative strategies that are characteristic for academic domains (Snow & Uccelli, 2009), as well as with language features that pertain to specific cognitive procedures, such as generalization or abstraction (cf. Feilke, 2012; Gottlieb & Ernst-Slavit, 2014; Heller & Morek, 2015; Nagy & Townsend, 2012). To convey these functions, it is characterized by a more sophisticated use of grammar and vocabulary than everyday language and requires the mastery of an extensive range of linguistic features (Scarcella & Rumberger, 2000). On the semantic level, academic language is characterized by functional academic vocabulary (e.g., *synthesize* and *denominator*) that students encounter at school but rarely in less formal social settings (Bailey & Butler, 2003; Uesseler, Runge, & Redder, 2013). English academic language has also been described as lexically dense and, on the syntactic level, comprising verb forms in passive voice, long and complex noun, and prepositional phrases, nominalizations, and participial modifiers (Abedi & Lord, 2001; Schleppgrell, 2004).

During the last few years, a number of authors have also proposed taxonomies of academic language in German (e.g., Feilke, 2012; Gogolin & Lange, 2011; Uesseler et al., 2013; Vollmer & Thürmann, 2010). Morek and Heller (2012) offer a synopsis of the lexical, grammatical, and discourse features that have been described in these taxonomies. According to this conceptualization, the academic register is characterized by the features described in Table 1. For the most part, these features correspond to the language features that have been suggested as characteristics of academic language in English (e.g., Bailey, Huang, Shin, Farnsworth, & Butler, 2007; Snow & Uccelli, 2009).

Academic language in German and English is marked by a frequent use of the features described above, whereas these features occur to a lesser extent in everyday language (cf. Bailey, 2007; Schleppgrell, 2004). Accordingly, several researchers have argued that texts may be located on a continuum, with texts containing everyday language and academic language features at the end points (Berendes, Dragon, Weinert, Heppt, & Stanat, 2013; Snow, 2010). However, to date it remains unclear whether understanding listening texts that contain predominantly academic versus everyday language features is associated with different cognitive processes and hence with different facets of listening comprehension.

Table 1. *Features of academic language*

Lexical features
• Quality of lexis: diverse, subject specific
Prefix verbs (e.g., <i>to reverse, to preempt, to substitute</i>)
Nominal compounds (e.g., <i>bar graph, two-digit number, bottom line</i>)
Standardized technical terms (e.g., <i>rectangular, rule of three</i>)
• Lexical density
Content words instead of pronouns
Nominalizations and elaborate noun phrases (e.g., <i>legilization, editing, average breath-holding capacity</i>)
Syntactic features
• Sentences instead of prosodic segmentation
• Local coherence by
Cohesion markers (e.g., conjunctions)
Complex sentences (e.g., relative, conjunctive, and disjunctive clauses; infinitival, participle clauses)
• Mode of representation
Declarative mood
Impersonal expressions (e.g., agentless passives)
Discursive features
• Speaker roles and turn taking organization (pre)determined
• Monological forms (e.g., lecture, presentation, essay)
• Subject-specific text types (e.g., minutes, report)
• Stylistic standards (e.g., objectivity, well structured, adequate length of text)

Note: Adapted from “Academic Discourse as Situated Practice: An Introduction,” by V. Heller and M. Morek, 2015, *Linguistics and Education*, 31, p. 176. Copyright 2015 by Elsevier. Adapted with permission.

ACADEMIC LANGUAGE IN SECOND LANGUAGE (L2) LEARNERS

L2 and first language (L1) students both need to acquire academic language when they enter school because it forms the basis of the acquisition of school-related skills. However, L2 children may face a double challenge: they need to learn the school-specific register, and they need to do so in a language that is not their native one (Aarts, Demir, & Vallen, 2011). Some researchers assume that L2 speakers acquire conversational language faster than academic language (Collier, 1989; Roessingh, Kover, & Watt, 2005). Cummins (2008), for example, reports a gap of several years between the attainment of conversational fluency, which is usually acquired within 2 years of exposure to a particular language, and the attainment of academic language proficiency, which may take 5 to 7 years. Analyzing data from four different school districts in the United States, Hakuta, Butler, and Witt (2000) confirmed this assumption and showed that it took English language learners 3 to 5 years to develop everyday language skills in English while the acquisition of academic language proficiency in English took 4 to 7 years. The rate of academic language attainment in L2 may depend on the age of arrival and prior schooling in L1: older children from ages 8 to 12 and adolescents who have already completed

several years of schooling in their L1 tend to acquire L2 academic language faster than younger children who did not attend school in their L1 (Collier, 1989). However, even the academic language skills of advanced L2 learners may not reach those of L1 students (Mason, 1995).

CRITICISM OF AND EMPIRICAL RESULTS ON ACADEMIC LANGUAGE

The differentiation between BICS and CALP proposed by Cummins (2000) has been criticized for a variety of reasons (for an overview, see Cummins, 2008). The concept might be of limited practical utility (Scarcella, 2003) because it does not specify which language skills are encompassed in CALP (Snow & Uccelli, 2009) and thus may not suffice to construct tasks and promote learning. Given the absence of a detailed description of CALP skills, Edelsky et al. (1983) propose that academic language is an artifact of the way it is measured and hence represents nothing but a form of “testwiseness.” Furthermore, it is questionable whether everyday language is always contextualized and less cognitively demanding than academic language. Persuasion, deceiving, or winning over others in everyday life, for example, requires cognitive abilities as complex as those needed to comprehend a persuasive argument in a social studies textbook (Bailey & Huang, 2011).

To date, only few studies have sought to differentiate academic language and everyday language empirically. Evidence concerning the assumed differences between everyday conversational listening and academic listening is particularly scarce (Buck, 2001). In Germany, first studies are presently carried out to identify German academic language features on an empirical basis. Most of these studies analyze language features occurring in German academic textbooks in the upper grades (Ahrenholz, 2013; Ahrenholz & Maak, 2012; Meurers, Berendes, Vajjala, & Bryant, 2015). Furthermore, a current research project aims at establishing a corpus of the German language used in schools. This corpus comprises transcripts of oral communication in German classrooms as well as textbooks and written texts by students in the primary grades (Köhne, Kronenwerth, Redder, Schuth, & Weinert, 2015). Although conclusive results from these projects are not yet available, preliminary findings indicate that the features discussed as characteristics of German academic language on a theoretical basis do occur in classroom discourse and textbooks.

Research that focused on written English language has also provided some empirical support for the assumption that academic and everyday language might be separable. Schleppegrell (2004) analyzed linguistic features of academic texts and transcripts of everyday conversations and found evidence for the theoretically postulated differences between academic and everyday language. In a similar vein, a number of studies indicated that the linguistic complexity of test items in mathematics or science tests impacted item difficulties differentially (e.g., Martiniello, 2008; Shaftel, Belton-Kocher, Glasnapp, & Poggio, 2006). The specific difficulty of test items containing academic language features compared to items containing everyday language features provides some support for the assumption of different language registers.

For listening comprehension of academic language, Shohamy and Inbar (1991) investigated the effects of different types of listening texts (oral oriented vs. literal

oriented) on L2 students in 12th grade. The authors did not explicitly mention that they were measuring academic and everyday listening comprehension, but the operationalization of the oral/literate listening texts is in line with central assumptions on academic and everyday listening. The results indicate that listening texts containing academic language features were significantly more difficult to process for the L2 listeners than everyday listening texts, thus supporting the notion that academic listening poses special demands on L2 listeners. Because Shohamy and Inbar (1991) did not include an L1 comparison group in their study, however, it remains unclear whether differences in academic versus everyday listening comprehension may also be found in native students and whether L2 listeners are differentially affected by academic language. Furthermore, the main focus of this study was on oral versus literate texts, such that the authors did not include all features that may be relevant to differentiate academic from everyday listening.

An initial study in Germany offers empirical support for the differentiation of German academic versus everyday listening in terms of the underlying morphosyntactic and lexical features (Gantefort & Roth, 2010; Roth, Neumann, & Gogolin, 2007). In this study with bilingual elementary students in Grade 4, oral classroom discourse was videotaped, transliterated, and then analyzed using confirmatory factor analysis. The authors thus were able to identify different modes of oral language within the classrooms: First, a *conversational mode*, in which students used colloquial language, such as high-frequency words or colloquial expressions, and second, an *academic mode*, which was marked by a high frequency of nominalizations, compounds, clause connectors, and impersonal expressions. Third, the authors identified an *elaborated mode*, which was characterized by the frequent use of subjunctives and constructions in passive voice. The authors postulate that the main features of German academic language are those typical for the academic and the elaborated mode (cf. Gantefort & Roth, 2010). However, it must be noted that no transcripts of conversational language outside the classroom were made, thus leaving unclear whether the conversational mode adequately reflects the conversational language students use in their everyday lives.

In line with theories on academic language in L2 students, some studies with English language learners have lent support to the notion that academic language may be especially difficult to acquire for these students. Uccelli, Galloway, Barr, Meneeses, and Dobbs (2015), for example, investigated academic language proficiency in elementary school students. Their operationalization of academic language skills not only pertains to the linguistic features of the register but also offers a more integrated conceptualization of academic language, including further skills such as connecting ideas logically or tracking participants and themes. The results from this study indicate that L2 students performed more poorly than L1 students on the academic language test. Similarly, Townsend, Filippini, Collins, and Biancarosa (2012) showed that L1 middle school students outperformed their L2 counterparts on a measure of academic vocabulary. It must be noted that both studies focused on academic language competencies, but did not investigate students' everyday language competencies. Hence, these studies do not provide evidence that differences in L1 and L2 students' academic language competencies might be more pronounced than differences in their everyday language competencies.

A number of other studies have challenged the assumption that academic language is particularly difficult for L2 students (Eckhardt, 2008; Heppt, Stanat, Dragon, Berendes, & Weinert, 2014). Eckhardt (2008) reported for German elementary students that L1 and L2 students performed less well on listening comprehension tasks that contained complex grammatical features and academic vocabulary. Furthermore, she found weaker performance for both groups of students on texts with school-related contents as opposed to texts with everyday contents. In a third study, Eckhardt (2008) varied the contextual embedding of the listening comprehension tasks. Texts that were context reduced (verbally described film scenes) were clearly more difficult to understand for elementary school students than texts that were supported by visual context (acted film scenes). Contrary to the author's expectation, these characteristics did not generally affect the performance of L2 students more than the performance of L1 students. Only for texts that contained school-related content, Eckhardt (2008) found L2 students' performance to be more strongly affected than L1 students' performance, but this effect disappeared when students' socioeconomic background was controlled.

This finding adds to a number of studies indicating that children from different socioeconomic backgrounds differ considerably in terms of language proficiency (Hoff, 2013). Aarts et al. (2011), for example, point out that mothers from families of high socioeconomic status (SES) tend to use more academic language features when talking to their children than mothers from families of lower SES. Given that L2 students are often disadvantaged in terms of their socioeconomic background (Organization for Economic Co-operation and Development, 2010), it is necessary to disentangle effects of language background and SES when comparing academic language skills in L1 and L2 students.

Taken together, empirical research has supported the assumption that processing academic language can be more difficult than processing everyday language. Contrary to theoretical assumptions on academic language, however, studies have not consistently shown that L2 students are disproportionately disadvantaged in their comprehension of academic language when compared to L1 students. This may result from the fact that most studies have focused on beginning level L2 learners with limited language skills, or on students in elementary grades. Because academic language closely resembles everyday language in elementary grades, it may be assumed that disadvantages of L2 students with academic language texts do not become visible until the later grades, when language becomes increasingly demanding in terms of academic features (Chudaske, 2012).

Furthermore, previous research has focused on academic language in terms of classification of texts and tasks. Yet, academic language is associated with specific language characteristics, such as sophisticated grammatical and lexical features that are associated with specific cognitive processes and demand a more elaborated language proficiency than that needed to process everyday language (cf. Gottlieb & Ernst-Slavit, 2014; Scarcella & Rumberger, 2000). It may therefore be assumed that the processing of academic language constitutes a specific facet of language proficiency. It remains unclear, however, whether listening comprehension of academic language and everyday language are associated with distinct facets of language proficiency.

RESEARCH QUESTIONS AND HYPOTHESES

To date, the findings on academic language comprehension in older students are inconclusive and empirical evidence on the differentiation between academic and everyday language comprehension is scarce. In our study, we varied features of academic language and investigated whether listening comprehension of academic language and listening comprehension of everyday language may be separable on the basis of these features in a sample of adolescent L1 and L2 students. Because grammatical features and, even more so, vocabulary knowledge are the most important aspects of academic oral proficiency in learning another language (Nagy & Townsend, 2012; Saville-Troike, 1984), we based our study on the CRESST definition of academic language that puts a special emphasis on linguistic features of academic language. The following research questions and hypotheses are addressed:

1. Can listening comprehension of academic language and everyday language be differentiated on the basis of features of academic language suggested in the literature (school-based content, reduced context, linguistic complexity)? We expect that academic language comprehension and everyday language comprehension may be distinguished empirically and that they show differential correlations to students' vocabulary and syntactic knowledge. More specifically, we assume students' vocabulary and syntactic knowledge to be more closely associated with their academic language comprehension than with their everyday language comprehension because academic language is assumed to entail vocabulary and syntactic features that are more difficult to process.
2. Are there performance differences between adolescent L1 and L2 students in listening comprehension of academic and everyday language? We expect L2 students to generally perform lower than L1 students on measures of listening comprehension. Based on theoretical assumptions that assume academic language to be more challenging for L2 adolescents than for L1 adolescents, we predict that these performance differences are more pronounced for listening comprehension of academic language than for listening comprehension of everyday language.

METHOD

Participants

There were 479 students from 20 schools who took part in the study. Of these, 20 students did not complete any of the test items and were therefore excluded from the analyses, resulting in a total of 459 students (45% female, 55% male). All students were in ninth grade, and their mean age was 16.3 years ($SD = 0.9$). The majority of students (43%) attended the lowest secondary school track of the German educational system (*Hauptschule*), 25% attended the intermediate track (*Realschule*), and 32% attended a comprehensive school (*Gesamtschule*), which provides several courses of education under one roof. All schools were public schools with 60% or more students with an immigration background and were located in urban districts with a high proportion of low-income or immigrant background population. Students who reported to have learned German as their

L1 within their families were designated as the L1 group ($n = 173$). Students who reported a language other than German as their L1 were designated as the L2 group ($n = 286$). The most frequent L1s in the latter group were Turkish (70.0%), Arabic (7.2%), and Italian (6.7%). Although most L2 students were born in Germany (76.6%), the majority indicated speaking a language other than German in their homes (67.9%). Most of the L2 students (89.9%) started speaking German before the age of 6 years, such that they had been speaking German for at least 10 years. Another 3.2% of L2 students had acquired German for more than 7 years, and the remaining 6.9% of L2 students had started acquiring German after the age of 10 years. The latter students' teachers evaluated their German language skills as being sufficiently developed; hence, none of these students was excluded from the analyses. All L1 students were born in Germany and had parents who were born in Germany. L1 and L2 students of our sample were distributed across the educational tracks in a similar manner, $\chi^2(2) = 0.44, p = .80$. The language groups did not differ in terms of age, $F(1, 421) = 1.28, p = .26$, or gender, $\chi^2(1) = 0.10, p = .75$. Furthermore, we found no group differences in students' basic cognitive abilities, $F(1, 391) = 2.43, p = .12$, as measured with a test on figural analogies (Kognitiver Fähigkeitstest für 4. bis 12. Klassen, Revision [Test of Cognitive Abilities for Grades 4 to 12, Revised]; Heller & Perleth, 2000).

Instruments

Listening comprehension of academic and everyday language. To assess listening comprehension of academic and everyday language, we developed a computer-based test. Students listened to German texts with a length of 100 to 150 words varying between 42 and 65 s in duration. All listening passages were written for the test and were spoken by German native speakers. Four texts (50%) were dialogues between two persons (1 male, 1 female), two texts were narratives delivered by one speaker (25%), and two texts were expository texts (25%) delivered by one speaker. After listening to each of the eight texts, students answered 3 to 5 items, resulting in a total of 32 questions. Questions were either in multiple-choice format with four response options (19 items, 59%) or in constructed-response format requiring a short written response (13 items, 41%). Items either pertained to the texts' main idea, for example, deciding on where a conversation may have taken place (17 items, 53%), or aimed at local details included in the listening passage, for example, remembering a given definition (15 items, 47%). Students were given 1 point if they chose the correct answer from the multiple-choice questions and 0 points if they chose a wrong option. Constructed response items were coded following a coding scheme and were scored in the same way. The texts were only presented acoustically, while all test items were presented acoustically as well as in written form on the computer screen. As the focus of our study was on verbal aspects of academic language, we did not use visual input (e.g., videos) to accompany the listening texts.

The listening comprehension test consisted of two subtests: one was designed to capture listening comprehension of academic language and the other listening comprehension of everyday language. Both subtests included four listening texts with 16 items, respectively. Academic texts had contents pertaining to more formal

settings (e.g., a job interview) and were context reduced, that is, provided less information on the characters and the setting of the plot than the everyday texts. On the linguistic level, academic texts had more syntactic features that are typical for written discourse and have been discussed as being relevant for German academic language (cf. Morek & Heller, 2012), namely, passive voice, complex subordinate clauses, and complex noun phrases. Vocabulary was rated as academic following the CRESST coding scheme for academic vocabulary (Butler, Bailey, Stevens, & Huang, 2004). Two associate professors of German linguistics who were blind to the research questions and the designated classification of the texts rated the difficulty of the syntactic features, the vocabulary, and the context on a scale ranging from 1 = *very easy* to 6 = *very difficult*. Units with academic listening texts were rated to be more difficult in syntactic features ($M = 3.88$, $SD = 1.79$), vocabulary ($M = 3.25$, $SD = 0.71$), and content ($M = 3.25$, $SD = 0.50$) than units with everyday language ($M = 2.75$, $SD = 0.89$ for syntactic features; $M = 2.75$, $SD = 0.71$ for vocabulary; and $M = 1.5$, $SD = 1.03$ for content). Interrater reliability for the two raters was acceptable for syntactic features ($\kappa = 0.69$, $p < .01$), vocabulary ($\kappa = 0.70$, $p < .01$), and content ($\kappa = 0.66$, $p < .01$).

Besides the variation of syntax and content, academic listening texts were constructed such that they contained more academic vocabulary (e.g., *syllable* and *to curtail*) than everyday listening texts. To classify vocabulary as academic or everyday vocabulary, three raters classified each word in the listening texts following the CRESST criteria (Butler et al., 2004). Interrater reliability for the three raters was $\kappa = 0.66$ ($p < .01$). Appendix A provides an overview of the listening text features for academic and everyday language units. Sample texts and items can be found in Appendix B.

As described above, academic listening and everyday listening *texts* differed in terms of their linguistic features, their content, and their contextualization. However, *items* for academic units as well as those for everyday units were constructed such that they were comparable in terms of their linguistic difficulty. Academic language and everyday language units were comparable pertaining to the number of items in multiple-choice versus constructed-response format, as well as in terms of items on main ideas versus local details from the text. The tests' internal consistency was $\alpha = 0.71$ for academic listening comprehension and $\alpha = 0.59$ for everyday listening comprehension. The manifest correlation between the two subtests was $r = .54$ ($p < .01$).

Vocabulary. We assessed vocabulary using the Culture Fair Intelligence Test—Scale A (Weiß, 2006). The test consists of 30 multiple-choice items that require students to find a synonym for a given German word from a list of five words. Its internal consistency was $\alpha = 0.81$.

Syntactic knowledge. To assess students' syntactic knowledge, we developed a computer-based test that consists of written sentences with various syntactical structures. Students had to decide for each sentence whether it was syntactically correct or not. To control for guessing, students could choose between the answers *incorrect*, *correct*, and *unclear*, with the latter option indicating that they were uncertain in their judgment (sample item: *Is the following sentence grammatically*

correct or incorrect? “*I to bed went late yesterday.*” *correct—incorrect—unclear*). The *unclear* option was later coded as wrong. The test includes 22 items and had an internal consistency of $\alpha = 0.86$ in the study sample.

SES. We used the Index of Economic, Social, and Cultural Status (ESCS; Organization for Economic Co-operation and Development, 2010) as an indicator of students’ SES. This measure is also employed in the Programme for International Student Assessment (PISA) studies and has a mean of 0 and a standard deviation of 1 in the PISA student population. The components comprising the ESCS are the families’ home possessions (such as books at home, cars, and televisions), the highest parental occupation as measured by the International Socioeconomic Index of Occupational Status (Ganzeboom, De Graaf, & Treiman, 1992), and the highest parental education in years of schooling. When levels of parental occupation or education differ for mother and father, the highest level is used to indicate a families’ occupational and educational status, respectively. All data were obtained with student and parent questionnaires.

Analyses

We first tested the measures of listening comprehension, vocabulary, and syntactic knowledge for items displaying differential item functioning (DIF) as a function of language background. DIF occurs when students with comparable abilities but from different groups have different probabilities of success on an item (Clauser & Mazor, 1998). Item DIF can be adverse or benign (e.g., Gierl, 2005). Benign item DIF is typically due to a multidimensionality of the construct and should therefore not be eliminated from the test. Adverse DIF can be attributed to the presence of nuisance dimensions and therefore threatens construct validity, such that items displaying adverse DIF between L1 and L2 students should be omitted.

DIF for each item was operationalized as the difference of item logit parameters between L1 and L2 students. The values were judged not only in terms of statistical significance but also in terms of effect size. Logit differences above 0.64 are considered as large magnitude DIF (cf. Wilson, 2005, p. 167). DIF, which persists in a two-dimensional Rasch model of listening comprehension, is considered to be adverse. Items that displayed adverse and large magnitude DIF for the language groups were excluded from the following analyses. The DIF analyses were conducted using the computer program ConQuest 2.0 (Wu, Adams, Wilson, & Haldane, 2007).

To test the first hypothesis on the separability of everyday listening and academic listening comprehension, we conducted a confirmatory factor analysis using Mplus 7.11 (Muthén & Muthén, 1998–2012). Students’ ESCS score was entered as a manifest covariate into the analyses. We estimated a two-dimensional model with items loading either on an academic or on an everyday language factor and then compared this model to a unidimensional model with a global factor of listening comprehension. Parameter estimates for the models are based on the robust weighted least square estimator for categorical outcomes (Muthén & Muthén, 1998–2012). We used the comparative fit index, the Tucker–Lewis index, and the root mean square error of approximation to evaluate goodness of fit.

Models with a Tucker–Lewis index and comparative fit index coefficient higher than 0.95 and a root mean square error of approximation coefficient smaller than 0.06 are regarded as good approximations to the data (Hu & Bentler, 1999). The models were then compared using chi-square difference tests. A significant chi-square difference of nested models suggests that the less restrictive model fits the data better than the more restrictive model.

To further investigate dimensionality, we estimated correlations between listening comprehension of academic and everyday language and receptive vocabulary and syntactic knowledge within the models. As listening comprehension of academic language is more complex than listening comprehension of everyday language in terms of vocabulary and syntax, listening comprehension of academic language should be more strongly associated with vocabulary and syntax than listening comprehension of everyday language. Both external criteria were modeled as unidimensional latent variables. We used Wald chi-square tests to determine differences in the correlations between the external criteria and listening comprehension of academic and everyday language (Wald, 1943).

In order to compare latent variable means between L1 and L2 students, as mentioned in our second hypothesis, we first tested the resulting listening comprehension model for measurement invariance to ensure applicability of the model to both groups of students (Baumgartner & Steenkamp, 1998). We initially estimated a baseline model, with the factorial structure set to be equal for the language groups but factor loadings and means allowed to vary (*configural invariance*). Subsequently, factor loadings (*metric invariance*) and, finally, means were restricted to be equal across groups as well (*scalar invariance*). Again, we evaluated goodness of fit according to model fit parameters and chi-square difference tests. We then compared latent variable means between L1 and L2 students using Wald chi-square tests (Wald, 1943).

RESULTS

Descriptive results

Table 2 provides the means, standard deviations, and intercorrelations for all measures. The results indicate that the majority of students in our sample grew up in low-SES families. The samples' ESCS mean was -0.63 ($SD = 0.92$), which is more than 0.5 SD below the German average (Organization for Economic Cooperation and Development, 2010). The language groups differed in terms of their socioeconomic background, with L2 students coming from families with lower SES than L1 students ($M = -0.33$, $SD = 0.84$ for L1 students; $M = -0.79$, $SD = 0.94$ for L2 students), $F(1, 363) = 17.50$, $p < .01$. Table 2 also shows that ESCS is only correlated with syntactic knowledge, vocabulary, and academic listening comprehension, but not with everyday listening comprehension. Therefore, it is appropriate to include ESCS as a covariate.

We did not find any language-group differences in the syntactic knowledge measure, which appeared to be relatively easy for all students. The L1 and L2 students solved 88% and 84% of the tasks, respectively, $F(1, 382) = 3.21$, $p = .07$. Although students also performed well on the vocabulary test, L2 students

Table 2. Means and standard deviations of all measures

	<i>N</i>	<i>M</i>	<i>SD</i>	Intercorrelations ^a				
				2	3	4	5	6
1. ESCS	363	−0.63	0.92	.26*	.16*	.16*	.09	.20*
2. Vocabulary ^b	392	0.75	0.15		.33*	.44*	.33*	.43*
3. Syntactic knowledge ^b	382	0.84	0.18			.45*	.42*	.37*
4. Listening comprehension (total) ^b	417	0.61	0.17				.85*	.90*
5. Listening comprehension (EL) ^b	417	0.68	0.18					.54*
6. Listening comprehension (AL) ^b	417	0.54	0.20					

Note: ESCS, Index of Economic, Social, and Cultural Status, reported as weighted likelihood estimates. EL, Everyday language; AL, academic language.

^aZero-order correlation coefficients with sample sizes of 357 ≤ *N* ≤ 417.

^bReported as percentage of correct answers.

**p* < .01.

showed significantly weaker vocabulary skills than L1 students (*M* = 0.80, *SD* = 0.12 for L1 students and *M* = 0.72, *SD* = 0.16 for L2 students), *F* (1, 392) = 14.30, *p* < .01. L2 students also performed less well than L1 students on the listening comprehension test in general (*M* = 0.66, *SD* = 0.14 for L1 students and *M* = 0.58, *SD* = 0.17 for L2 students), *F* (1, 417) = 29.48, *p* < .01. Both groups of students showed weaker performance on the academic listening subtest (*M* = 0.60, *SD* = 0.18 for L1 students and *M* = 0.50, *SD* = 0.21 for L2 students) than on the everyday listening subtest (*M* = 0.73, *SD* = 0.15 for L1 students and *M* = 0.65, *SD* = 0.18 for L2 students), yet the difference was not significant within either of the groups, *t* (153) = −0.86, *p* = .93 for L1 students and *t* (262) = 0.69, *p* = .95 for L2 students.

Detecting DIF

The results of the DIF analyses indicated that the listening comprehension test as well as the vocabulary test and the test of syntactic knowledge contained several items displaying significant DIF as a function of language background. We identified three items (two everyday listening items and one academic listening item) as displaying adverse and large magnitude DIF for the listening comprehension test (9.4%), seven items (23.3%) for the vocabulary test, and two items (9.1%) for the test of syntactic knowledge. These items were excluded from further analyses.

Dimensionality of listening comprehension

Table 3 shows the data fit for the unidimensional and the two-dimensional model of listening comprehension. The fit indices suggest that both models fit the data well. However, the comparison of the models using chi-square difference testing indicates that the fit differs. The less restrictive two-dimensional model fits the data significantly better than the more restrictive unidimensional model. Furthermore,

Table 3. *Test of dimensionality for the listening comprehension test (controlling for ESCS)*

Model	χ^2	df	p	RMSEA	CFI	TLI	Compared to	$\Delta\chi^2$	df	p
A: unidimensional	492.04	405	<.01	0.02	0.94	0.94	—	—	—	—
B: two dimensional	473.49	403	.01	0.02	0.95	0.95	Model A	18.15	2	<.01

Note: ESCS, Index of Economic, Social, and Cultural Status; RMSEA, root mean square error of approximation; CFI, comparative fit index; TLI, Tucker–Lewis index; $N = 459$.

the fit indices comparative fit index and Tucker–Lewis index improved in the two-dimensional model, thus providing further support for the assumption that the two-dimensional model better fits the data. At the same time, listening comprehension for academic language and listening comprehension for everyday language are highly correlated ($r = .84, p < .01$), suggesting substantial overlap between the two facets. We therefore conducted a post hoc test to analyze whether this correlation is significantly lower than one. For this purpose, we fixed the correlation between the two facets to one and compared this model with a model where the correlation was not restricted. The nested model test indicates that the correlation between the latent variables is significantly different from 1 ($\Delta\chi^2 = 11.92, p < .01$).

We further examined the dimensionality by estimating the correlations between the two listening comprehension scores and vocabulary and syntactic knowledge as external criteria. Figure 1 shows the concurrent correlations between listening comprehension of academic and everyday language on the one hand and students’ receptive vocabulary and syntactic knowledge on the other hand, controlling for ESCS. The results indicate that syntactic knowledge is more strongly associated with listening comprehension of everyday language than with listening comprehension of academic language ($\chi^2 = 11.33, p = .01$). Receptive vocabulary, in contrast, correlates somewhat more strongly with listening comprehension of academic language than with listening comprehension of everyday language, yet the difference is not significant ($\chi^2 = 0.38, p = .54$). Thus, we did not find that vocabulary and syntactic knowledge were more strongly associated with academic language comprehension as suggested by our first hypothesis. The proportion of explained variance was $R^2 = .46$ for listening comprehension of everyday language and $R^2 = .49$ for listening comprehension of academic language.

Group differences in listening comprehension of academic language and listening comprehension of everyday language

In order to test our second hypothesis on group differences between L1 and L2 students in terms of listening comprehension of everyday and academic language, we examined whether measurement invariance is given for the two groups and, thus, whether latent means can be compared. As goodness of fit does not significantly decrease in the more restrictive models (Table 4), scalar measurement invariance can be assumed and means can be reasonably compared across groups.

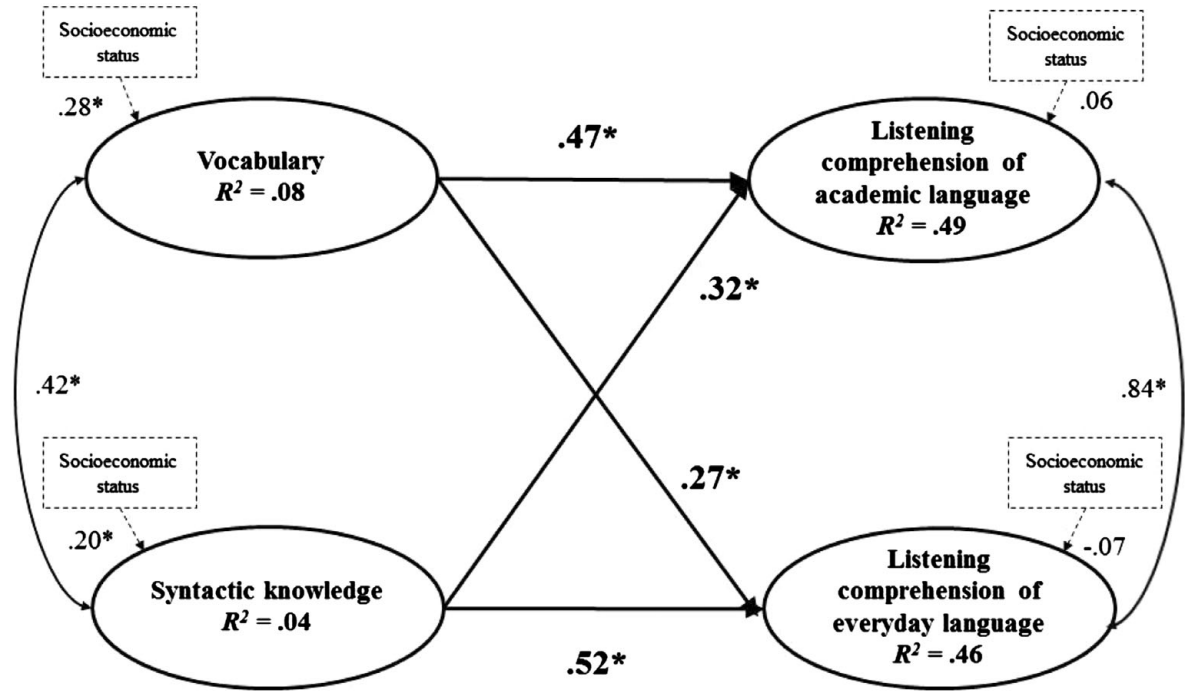


Figure 1. Correlations between listening comprehension of academic and everyday language vocabulary and syntactic knowledge. Indicators are not depicted. $\chi^2 = 2,720.58$, $df = 2,546$, $p < .01$, root mean square error of approximation > 0.01 , comparative fit index = 0.95, Tucker–Lewis index = 0.95 ($N = 459$).

Table 4. *Test of measurement invariance for the two-dimensional model (controlling for ESCS)*

Model	χ^2	df	p	RMSEA	CFI	TLI	Compared to	$\Delta \chi^2$	df	p
Configural	859.73	806	.09	0.02	0.95	0.94	—	—	—	—
Metric	870.46	833	.18	0.01	0.96	0.96	Configural	21.52	27	.76
Scalar	897.40	860	.18	0.01	0.96	0.96	Metric	28.00	27	.41

Note: ESCS, Index of Economic, Social, and Cultural Status; RMSEA, root mean square error of approximation; CFI, comparative fit index; TLI, Tucker–Lewis index; $N = 459$.

Table 5. *Variable means for L1 and L2 students*

	L1 ($n = 173$)		L2 ($n = 286$)		Difference Testing for Latent Variable Means		
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	Wald χ^2	<i>p</i>	<i>d</i> ^a
Listening comprehension (EL)	0.73	0.15	0.65	0.18	12.98	<.01	−0.65
Listening comprehension (AL)	0.60	0.18	0.50	0.21	4.68	.03	−0.45

Note. L1, First language; L2, second language; EL, everyday language; AL, academic language.

^aDifference between latent means of the L1 group (reference group) and the L2 group.

We then tested our second hypothesis on differences in listening comprehension of academic and everyday language scores between L1 and L2 students by comparing latent variable means between the two language groups using Wald χ^2 tests (Wald, 1943). Latent variable means of L2 students were compared to the reference group of L1 students for which the latent means were fixed to zero. The results in Table 5 indicate that L2 students performed less well than L1 students on both facets of listening comprehension. Contrary to what we expected, differences between L1 and L2 students’ listening comprehension of everyday language were more pronounced than differences between listening comprehension of academic language ($\chi^2 = 8.72, p < .01$).

DISCUSSION

The study investigated listening comprehension of academic and everyday language in students speaking German as their L1 or as their L2. We hypothesized that academic language and everyday language form separate facets of listening comprehension. For both groups of students, we assumed that listening comprehension of academic language would be more difficult than listening comprehension of everyday language. We also expected that L2 students would generally perform less well than L1 students, yet that the performance gap would be more pronounced for listening comprehension of academic language than for listening comprehension of everyday language.

Dimensionality of listening comprehension in L1 and L2 learners

Concerning the first research question on the structure of listening comprehension, findings of a confirmatory factor analysis indicated that a two-dimensional model fits the data better than a unidimensional model. Our results thus provide preliminary evidence that listening comprehension of academic and everyday language may be distinguished from each other empirically. This differentiation is in line with theoretical assumptions by Richards (1983), who postulated that conversational listening and academic listening rely on different skills as well as with conceptualizations that suppose academic language to be associated with the mastery of specific genres, strategies, and linguistic features (Gottlieb & Ernst-Slavit, 2014; Nagy & Townsend, 2012; Scarcella & Rumberger, 2000; Snow & Uccelli, 2009). As Heller and Morek (2015) propose, the register of academic language has specific functions, such as transmission of knowledge or acting as a tool for thinking. While the first function is closely related to defining academic language as a register that serves specific communicative demands (e.g., producing or understanding decontextualized and explicitly structured complex information), the latter function of academic language pertains to qualitatively different ways of thinking as compared to everyday language: students who use academic language correctly are supposed to be capable of executing the complex cognitive operations (e.g., abstraction and generalization) that are central to scientific thinking (Morek & Heller, 2012; Vollmer & Thürmann, 2010). The two-dimensional structure of listening comprehension is in line with this notion of qualitatively different cognitive processes that are closely associated with the use of academic features.

It must be noted that our study does not allow us to differentiate whether our two-dimensional model reflects different aspects of language comprehension or differences in listening texts' difficulty. However, as academic language is by definition more complex and difficult to process than everyday language, a greater difficulty of academic language demands is inherent to the construct and is hence also reflected in our operationalization of academic listening comprehension. Our findings thus may also be seen to corroborate results by Shohamy and Inbar (1991), who found that the amount of oral language features in listening texts had an impact on L2 test takers' listening comprehension. The present study adds to this research and indicates not only that differences in the processing of academic versus everyday listening texts may be a matter of text type and difficulty but also that listening texts that are located near the extremes of the language continuum from academic language to everyday language may also be associated with different facets of listening comprehension.

Although the results from the confirmatory factor analysis lent support for the proposed two-dimensionality of listening comprehension, an additional validation procedure using students' vocabulary and their syntactic knowledge as external criteria did not clearly support our assumption. Contrary to our predictions, students' listening comprehension of everyday language correlated more strongly with their syntactic knowledge than students' listening comprehension of academic language. As hypothesized, vocabulary correlated somewhat stronger with listening comprehension of academic language than with listening comprehension of everyday language. However, these differences were not significantly different.

The lack of expected associations between the vocabulary measure and listening comprehension of academic language might be due to our operationalization of students' vocabulary because this measure (Culture Fair Intelligence Test—Scale A; Weiß, 2006) does not provide a differentiation between academic and everyday vocabulary.

However, it is possible to classify the target words used in the test as either academic vocabulary or everyday vocabulary. Based on the coding scheme by the CRESST (Butler et al., 2004), we classified 7 (30.4%) words as academic vocabulary and 16 words (69.6%) as everyday vocabulary ($\kappa = 0.66$, $p < .01$ for three raters). The vocabulary measure thus entails words from both facets of language comprehension, which may explain the lack of differences in correlations between the vocabulary test and listening comprehension of academic and everyday language, respectively. This assumption is supported by the fact that the elimination of items containing everyday language from the vocabulary measure resulted in a significant correlation between vocabulary and academic listening comprehension ($r = .30$, $p = .02$) but a nonsignificant association between vocabulary and everyday language comprehension ($r = .16$, $p = .23$). The difference between the correlation coefficients is significant ($z = 2.24$, $p = .03$). Furthermore, the syntactic knowledge measure only included everyday language. The higher correlation between syntactic knowledge and everyday listening as compared to the correlation between syntactic knowledge and academic listening may thus also result from the operationalization.

Taken together, these results suggest that listening comprehension of academic texts might be differentiated from listening comprehension of everyday texts, although both facets of listening comprehension are closely associated. Additional analyses that only included the academic vocabulary items of our general vocabulary measure also indicate that listening comprehension of academic language might pose special demands on students' vocabulary. Our findings are thus not in line with assumptions by Scarcella (2003), who has criticized the dichotomous conceptualization of academic versus everyday language as not being "useful for understanding the complexities of academic English or the multiple variables affecting its development" (p. 5). According to her view, both academic language and everyday language are complex registers that share a number of important determinants (e.g., processing strategies). Against this background, Scarcella (2003) notes that the definition of academic language as being per se more complex and more difficult to process than everyday language might not be appropriate. Rather, both registers might exhibit a range of overlaps, and certain aspects of academic language might be acquired prior to the acquisition of some everyday language features. However, our findings provide preliminary evidence that a distinction between listening comprehension of everyday language and listening comprehension of academic language seems warranted.

Differences in academic and everyday listening comprehension between L1 and L2 learners

Concerning our second hypothesis on group differences in latent variable means for listening comprehension of academic and everyday language, our data indicated

that both groups of students showed weaker performance on the test of academic listening comprehension than on the test of everyday listening comprehension, and L1 students generally outperformed their L2 peers. However, our results failed to support the assumption that academic listening comprehension is especially demanding for L2 students. We thus could not find support for a central assumption by Cummins (1980) and other researchers (cf. Collier, 1989; Roessingh et al., 2005), namely, that academic language is disproportionately more difficult for L2 students than for L1 students. These findings are in line with results from previous studies with elementary school students in Germany that also did not find L2 students to be especially disadvantaged in terms of academic listening comprehension (Eckhardt, 2008; Heppt et al., 2014).

Contrary to what was expected, differences between L1 students' and L2 students' everyday listening comprehension were even more pronounced than differences in their academic listening comprehension. A possible explanation for this finding pertains to the restricted variance in students' socioeconomic background. The majority of students in our sample were from low-income and low-SES families. As research has indicated, the attainment of academic language is closely associated with students' socioeconomic background (Aarts et al., 2011; Hoff, 2013). Students who rarely encounter an "academic culture" of literacy (Corson, 1997, p. 672) outside school are thus less likely to develop academic language competencies than students from more privileged families. It is possible that low-SES L1 students in our study did not have sufficient opportunities to acquire academic language and were therefore also compromised in their academic language skills.

In a similar vein, the sociocultural component of academic language, that is, its function as instantiating a social identity, may have influenced the attainment of academic language in our low-SES sample (cf. Heller & Morek, 2015; Uccelli, Barr, et al., 2015). As Galloway, Stude, and Uccelli (2015) have shown for monolingual students from low-income families, middle school students reported to be aware of the academic register and its features; however, they mainly regarded it as a "vehicle for social positioning in the accomplishment of social situations" (Heller & Morek, 2015, cited after Galloway et al., 2015) and thus not necessarily evaluated it as a tool for communicating in classrooms. Because our sample also consisted of students from low-income families, it may be assumed that both L1 and L2 students did not perceive academic language primarily as a means for communicating efficiently and consequently did not make strong efforts to attain this register.

The more pronounced differences in listening comprehension of everyday language as compared to listening comprehension of academic language may also result from the fact that linguistic competencies are, to some extent, language specific and are acquired in either the L1 or the L2, depending on the time-on-task (Gathercole, 2002). As students from low-SES families rarely encounter academic language features outside the classroom (Reese, Garnier, Gallimore, & Goldenberg, 2000), both L1 and L2 students acquire this register within their schools and hence in the language of instruction. Even if academic language is rarely conveyed explicitly by teachers (cf. Corson, 1997), L1 and L2 students thus share the same learning opportunities to gain proficiency in German academic language.

At the same time, most of the L2 students in our study speak their respective L1 in their everyday life and have less contact to the register of German everyday language than their L1 peers. It can therefore be reasonably assumed that the pronounced differences between L1 and L2 students in listening comprehension of everyday language may result from the differential opportunities to acquire the conversational register in German.

Although our study does not allow us to test these assumptions, it may be surmised that the social context in which listening comprehension is acquired shapes its attainment in L1 and L2 students through the allocation of specific learning opportunities. As a result, L1 and L2 students' competencies may vary for language features that L1 and L2 students learn in different contexts, but may be more comparable for language features for which students share the same context of acquisition and hence the same time on task. In terms of the acquisition of general language features, empirical research has already demonstrated that L1 and L2 students from similar backgrounds exhibit comparable levels of language competencies (e.g., Kieffer, 2010). As our results suggest, the attainment of listening comprehension of academic language in L1 and L2 students may also be shaped, at least partially, by the context of acquisition.

Limitations

The present study has some limitations that have to be noted. Our sample was restricted to low-SES students from the lowest tracks of the educational system. Further studies are needed that investigate whether listening comprehension of academic and everyday language are more clearly distinguishable in students who have had more opportunities to gain proficiency in the register of academic language. Furthermore, the aim of the present study was to compare listening comprehension of academic and everyday language with a special emphasis on lexical and grammatical features of academic language that constitute this register (cf. Butler et al., 2004). However, academic language not only is characterized by linguistic features but also includes sociocultural aspects, such as knowledge about social and cultural norms, values, and habits that shape the context of academic language (Heller & Morek, 2015; Scarcella, 2003). These aspects were not considered in our operationalization of academic language. Therefore, internal validity is ensured, but ecological and external validity may be limited.

Although our results seem to support the assumption that academic and everyday registers may be regarded as different aspects of language comprehension, it must be noted that our structural perspective on academic language and its lexical, syntactic, and textual features may neglect other aspects of academic language that could lead to similar results as we observed. It was beyond the scope of our study to investigate the complex cognitive operations, such as abstraction, generalization, or logical argumentation, that are assumed to be closely intertwined with the exertion of linguistic academic language features. The present study therefore does not allow for responding to a main point of criticism on the construct of academic language, namely, that it is not necessarily more cognitively demanding than everyday language (cf. Bailey & Huang, 2011). While we found

support for the assumption that, from a structural perspective, academic and everyday listening comprehension may be regarded as different facets of listening comprehension, further research is needed that investigates the cognitive operations that are associated with these two registers. Future studies could build on recent work by Uccelli, Barr, et al. (2015), who have proposed a more comprehensive conceptualization of the language skills needed in academic contexts. The authors describe a set of core academic language skills, which are defined as “knowledge and deployment of a repertoire of language forms and functions that co-occur with oral and written school learning tasks across disciplines” (p. 1079). Further research is needed to investigate the acquisition of these skills as well as to explore the strategies students use and the cognitive processes they engage in when confronted with academic versus everyday listening tasks (e.g., using think-aloud procedures).

Furthermore, it is important to note that we did not investigate naturalistic academic discourse in our study, but modeled listening comprehension at the extremes of the language continuum from academic language to everyday language. It is conjecturable that listening texts that occur in actual classroom discourse contain a wider range of everyday and academic language features, thus resulting in listening demands that may be located somewhat more to the middle of the language continuum. Including such texts might lead to different results or even result in unidimensionality of listening comprehension.

Conclusion

Despite its limitations, the present study indicates that listening comprehension of academic and everyday language may describe separable facets of listening comprehension. However, as the external validation procedure did not fully substantiate this finding, and as we did not find the expected performance differences between L1 and L2 students, our results only partially support the notion that academic language and everyday language are associated with distinct facets of listening comprehension. Nonetheless, they may serve as a starting point to further examine the cognitive processing of listening comprehension of academic language and everyday language. Such studies should also aim at investigating academic and everyday listening competencies in other populations than the low-track, low-SES students we considered in our study. For example, international studies indicate that academic language is more prevalent in the higher tracks of the educational system, thus providing high-track students with more opportunities to gain proficiency in this language register (e.g., Harklau, 1994). Everyday and academic listening comprehension may therefore require more distinct cognitive processes in higher tracks, where the materials used are more demanding in terms of academic language features. Notwithstanding the need for further research, our results provide first evidence that the understanding of academic versus everyday listening texts, at least when located somewhere near the extremes of the language continuum from academic language to everyday language, may require somewhat different cognitive processes.

As L2 students performed significantly below their native peers not only when listening to academic language texts but also when processing everyday language,

our results further indicate that L2 instruction should focus on academic language *and* everyday language skills to help L2 students gain proficiency in all registers of their L2. In Germany, second-generation L2 students typically attend regular German classrooms. However, as our results suggest, simply exposing L2 students to an immersion context seems not to be sufficient for developing listening competencies that are comparable to those of their monolingual peers. Further research is needed to investigate which approaches can effectively help L2 students gain proficiency in all aspects of L2 listening comprehension.

APPENDIX A

Table A.1. *Variation of linguistic features for academic and everyday listening texts*

Feature	EL ^a	AL ^b
	<i>M (SD)</i>	<i>M (SD)</i>
1. Number of words per text	138.25 (21.78)	108.25 (14.08)
Lexical features		
2. Number of nominalizations per text	0.75 (0.96)	1.00 (0.00)
3. Number of academic vocabulary per text	2.50 (2.08)	8.50 (3.42)
Syntactic features		
4. Sentence length (words per sentence)	10.50 (4.68)	15.18 (2.13)
5. Number of complex subordinate clauses per text	0.75 (0.96)	1.00 (0.00)
6. Number of passive constructions per text	0.00 (0.00)	3.25 (2.87)
7. Number of structural words per text ^c	58.08 (2.82)	56.32 (6.29)
Mean number of items correct		
Total sample	0.70 (0.17)	0.53 (0.20)
L1 students	0.74 (0.15)	0.59 (0.19)
L2 students	0.68 (0.18)	0.49 (0.20)

Note: EL, Listening comprehension of everyday language; AL, listening comprehension of academic language

^aContents: going shopping, doing homework, getting a driver’s license, friendship.

^bContents: haikus, ocelots, doing a job interview, planning a hike.

^cStructural words are words that are used for grammatical purposes only (e.g., prepositions, conjunctions).

APPENDIX B

Haiku

Short poems have been known since ancient Japan, where they were called *haiku*. The society at the emperor's court developed them from a sort of tournament because they were frequently bored. Back then, the high society liked to compete in rhyming verses; however, strict rules had to be obeyed. With regard to content, for instance, a haiku had to portray an image from nature. Furthermore, it was allowed to include only a certain amount of syllables and always had to exhibit the same length. The haiku, therefore, was the shortest poetic form in the world. Even today lots of people engage in composing haikus in their leisure time. For the most part, the old rules were adopted, so that a lot of modern haikus appear as if they were written in ancient Japan. Thus, the haiku is often declared to be a timeless form of poetry.

Question 1

What do you need to make a good haiku?

- A You need to be good at writing.
- B You need to be good at listening.
- C You need to be good at competing.
- D You need to be good at drawing.

Code 1: A: You need to be good at writing.

Code 0: All other responses

Question 2

What is a haiku?

- A A Japanese game.
- B A short poem.
- C A name for ancient Japan.
- D A Japanese form of song.

Code 1: B: A short poem.

Code 0: All other responses

Question 3

In the text, something is said about why haikus are special. Which kind of "record" do they hold?

Code 1: Mentions the fact that haikus are the shortest poetic form in the world

- They are the shortest.
- They are very short.
- The shortest poetic form (of the world).

Code 0: Answer is insufficient or vague. Shows insufficient understanding of the text or gives an implausible or irrelevant answer.

- A high record.
- They have strict rules.
- They are a timeless form of poetry.

Shopping

Salesperson: May I help you?

Client: Oh ... well ... yes, you can. I am looking for a ... this magazine for teens. That has a lot on music inside.

Salesperson: "Melody," right?

Client: Yeah, exactly, that was it. Unfortunately, I forgot my shopping list at home, it was on it. Do you have it, "Melody"?

Salesperson: Look at your left, next to the box with the lighters.

Client: Where? I only see comics.

Salesperson: No, one box further to the left.

Client: Ah, now I got it! Usually, my daughter gets the money from me and then buys her stuff by herself, you know. That's why I'm not familiar here. But now she's broken her leg.

Question 1

What does the client want to buy?

- A A lighter
- B A newspaper
- C A comic
- D A magazine for teens

Code 1: D: A magazine for teens

Code 0: All other responses

Question 2

The client has problems finding what she is looking for. Why is that? Use the information you have heard to answer the question.

Code 1: Refers to the fact that the client is in the shop for the first time OR on the fact that the daughter is usually going shopping herself OR to the shopping list that has been forgotten

- Because, normally, the daughter goes shopping.
- She is in the shop for the first time and doesn't know where things are.
- She forgot her shopping list.
- Because she is not familiar with the shop.
- She's there for the first time.

Code 0: Answer is insufficient or vague. Shows insufficient understanding of the text or gives an implausible or irrelevant answer.

- The daughter has broken her leg.
- She is looking for a magazine.
- Because there are so many comics.
- Because it is somewhere else.

Question 3

In what kind of shop does this conversation may have taken place? Use the information you have heard to answer the question.

Code 1: Mentions at least one sort of shop that usually sells magazines, comics, and tobacco supplies.

- At a kiosk
- At a lottery shop
- At a newspapers stand
- At a supermarket

Code 0: Answer is insufficient or vague. Gives an implausible or irrelevant answer or mentions shops that do not sell magazines and tobacco supplies at the same time.

- A shop where comics and magazines are sold.
- In Germany.
- In a comic store.

Question 4

In the text, the clients' daughter is mentioned. What do you think she does?

- A She goes to the kindergarten.
- B She visits high school.
- C She's studying in another city.
- D She lives abroad.

Code 1: B: She visits high school.

Code 0: All other responses.

ACKNOWLEDGMENTS

The present study was supported by grants from the German Science Foundation (Deutsche Forschungsgemeinschaft STA 626/1-6 and SCHN 315/40-1).

REFERENCES

- Aarts, R., Demir, S., & Vallen, T. (2011). Characteristics of academic language register occurring in caretaker-child interaction: Development and validation of a coding scheme. *Language Learning*, 61, 1173–1221. doi:10.1111/j.1467-9922.2011.00664.x

- Abedi, J., & Lord, C. (2001). The language factor in mathematics tests. *Applied Measurement in Education*, 14, 219–234. doi:10.1207/S15324818AME1403_2
- Ahrenholz, B. (2013). Sprache im Fachunterricht untersuchen [Investigating language in content area classrooms]. In C. Röhner & B. Hövelbrinks (Eds.), *Fachbezogene Sprachförderung in Deutsch als Zweitsprache—Theoretische Konzepte und empirische Befunde zum Erwerb bildungssprachlicher Kompetenzen* (pp. 87–98). Weinheim: Beltz.
- Ahrenholz, B., & Maak, D. (2012). Sprachliche Anforderungen im Fachunterricht: Eine Skizze mit Beispielanalysen zum Passivgebrauch in Biologie [Linguistic demands in content area classrooms: A draft with exemplary analyses on the use of passive voice in biology]. In H. Roll & A. Schilling (Eds.), *Mehrsprachiges Handeln im Fokus von Linguistik und Didaktik* (pp. 135–152). Duisburg: UBRR Verlag.
- Bailey, A. L. (Ed.) (2007). *The language demands of school: Putting academic English to the test*. New Haven, CT: Yale University Press.
- Bailey, A. L., & Butler, F. (2003). *An evidentiary framework for operationalizing academic language for broad application to K–12 education: A design document*. Los Angeles: University of California.
- Bailey, A. L., & Huang, B. H. (2011). Do current English language development/proficiency standards reflect the English needed for success in school? *Language Testing*, 28, 343–365. doi:10.1177/0265532211404187
- Bailey, A. L., Huang, B. H., Shin, H. W., Farnsworth, T., & Butler, F. A. (2007). *Developing academic English language proficiency prototypes for 5th grade reading: Psychometric and linguistic profiles of tasks. CRESST report 727*. Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing.
- Baumgartner, H., & Steenkamp, J.-B. E. M. (1998). Multi-group latent variable models for varying numbers of items and factors with cross-national and longitudinal applications. *Marketing Letters*, 9, 21–35. doi:10.1023/A:1007911903032
- Berendes, K., Dragon, N., Weinert, S., Heppt, B., & Stanat, P. (2013). Hürde Bildungssprache? Eine Annäherung an das Konzept “Bildungssprache” unter Einbezug aktueller empirischer Forschungsergebnisse [Academic language as an obstacle? Approaching the concept of “academic language” considering recent empirical findings]. In A. Redder & S. Weinert (Eds.), *Sprachförderung und Sprachdiagnostik: Perspektiven aus Psychologie, Sprachwissenschaft und empirischer Bildungsforschung* (pp. 17–41). Münster: Waxmann.
- Biber, D. (1995). *Dimensions of register variation*. New York: Cambridge University Press.
- Buck, G. (2001). *Assessing listening*. Cambridge: Cambridge University Press.
- Butler, F. A., Bailey, A. L., Stevens, R., & Huang, B. H. (2004). *Academic English in fifth-grade mathematics, science, and social studies textbooks*. Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing.
- Chamot, A. U., & O’Malley, J. M. (1994). *The CALLA handbook: Implementing the cognitive academic language learning approach*. Reading, MA: Addison-Wesley.
- Chudaske, J. (2012). *Sprache, Migration und schulfachliche Leistung* [Language, migration, and school performance]. Wiesbaden: VS Verlag.
- Clauser, B. E., & Mazor, K. M. (1998). Using statistical procedures to identify differentially functioning test items. *Educational Measurement: Issues and Practice*, 17, 31–44.
- Collier, V. P. (1989). How long? A synthesis of research on academic achievement in a second language. *TESOL Quarterly*, 23, 509–531. doi:10.2307/3586923
- Corson, D. (1997). The learning and use of academic English words. *Language Learning*, 47, 671–718. doi:10.1111/0023-8333.00025
- Csomay, E. (2006). Academic talk in American university classrooms: Crossing the boundaries of oral-literate discourse. *Journal of English for Academic Purposes*, 5, 117–135. doi:10.1016/j.jeap.2006.02.001

- Cummins, J. (1980). The construct of language proficiency in bilingual education. In J. E. Alatis (Ed.), *Current issues in bilingual education* (pp. 81–103). Washington, DC: Georgetown University Press.
- Cummins, J. (2000). *Language, power, and pedagogy: Bilingual children in the crossfire*. Clevedon: Multilingual Matters.
- Cummins, J. (2008). BICS and CALP: Empirical and theoretical status of the distinction. In B. Street & N. H. Hornberger (Eds.), *Encyclopedia of language and education: Vol. 2. Literacy* (2nd ed., pp. 71–83). New York: Springer Science + Business Media LLC.
- Curenton, S. M., & Justice, L. M. (2004). African American and Caucasian preschoolers' use of decontextualized language: Literate language features in oral narratives. *Language, Speech, and Hearing Services in Schools*, 35, 240–253. doi:0161-1461/04/3503-0240
- Eckhardt, A. G. (2008). *Sprache als Barriere für den schulischen Erfolg: Potentielle Schwierigkeiten beim Erwerb schulbezogener Sprache für Kinder mit Migrationshintergrund* [Language-related barriers in educational success: Potential difficulties in the acquisition of academic language in immigrant students]. Münster: Waxmann.
- Edelsky, C., Hudelson, S., Flores, B., Barkin, F., Altwerger, B., & Jilbert, K. (1983). Semilingualism and language deficit. *Applied Linguistics*, 4, 1–22. doi:10.1093/applin/4.1.1
- Feilke, H. (2012). Bildungssprachliche Kompetenzen-fördern und entwickeln [Fostering and developing academic language competencies]. *Praxis Deutsch*, 233, 1–13.
- Ferris, D., & Tagg, T. (1996). Academic oral communication needs of EAP learners: What subject-matter instructors actually require. *TESOL Quarterly*, 30, 31–58. doi:10.2307/3587606
- Flowerdew, J. (1994). *Academic listening: Research perspectives*. Cambridge: Cambridge University Press.
- Flowerdew, J. (1995). Research of relevance to second language lecture comprehension: An overview. In J. Flowerdew (Ed.), *Academic listening: Research perspectives* (pp. 7–29). Cambridge: Cambridge University Press.
- Flowerdew, J., & Miller, L. (1997). The teaching of academic listening comprehension and the question of authenticity. *English for Specific Purposes*, 16, 27–46. doi:10.1016/S0889-4906(96)00030-0
- Galloway, E. P., Stude, J., & Uccelli, P. (2015). Adolescents' metalinguistic reflections on the academic register in speech and writing. *Linguistics and Education*, 31, 221–237. doi:10.1016/j.linged.2014.10.006
- Gantefort, C., & Roth, H.-J. (2010). Sprachdiagnostische Grundlagen für die Förderung bildungssprachlicher Fähigkeiten [Assessment of academic language proficiency in German]. *Zeitschrift für Erziehungswissenschaft*, 13, 573–591. doi:10.1007/s11618-010-0163-2
- Ganzeboom, H. B. G., De Graaf, P. M., & Treiman, D. J. (1992). A standard international socioeconomic index of occupational status. *Social Science Research*, 21, 1–56. doi:10.1016/0049-089X(92)90017-B
- Gathercole, V. C. M. (2002). Monolingual and bilingual acquisition: Learning different treatments of that-trace phenomena in English and Spanish. In D. K. Oller & R. E. Eilers (Eds.), *Language and literacy in bilingual children* (pp. 220–252). Clevedon: Multilingual Matters.
- Gierl, M. J. (2005). Using dimensionality-based DIF analyses to identify and interpret constructs that elicit group differences. *Educational Measurement: Issues and Practice*, 24, 3–14. doi:10.1111/j.1745-3992.2005.00002.x
- Gogolin, I., & Lange, I. (2011). Bildungssprache und Durchgängige Sprachbildung [Academic language and integrated language education]. In S. Fürstenau & M. Gomolla (Eds.), *Migration und schulischer Wandel: Mehrsprachigkeit* (pp. 107–128). Wiesbaden: VS Verlag.
- Gottlieb, M., & Ernst-Slavit, G. (2014). *Academic language in diverse classrooms: Definitions and contexts*. Thousand Oaks, CA: Corwin Press.
- Haag, N., Heppt, B., Stanat, P., Kuhl, P., & Pant, H. A. (2013). Second language learners' performance in mathematics: Disentangling the effects of academic language features. *Learning and Instruction*, 28, 24–34. doi:10.1016/j.learninstruc.2013.04.001

- Hakuta, K., Butler, Y. G., & Witt, D. (2000). *How long does it take English learners to attain proficiency?* Santa Barbara, CA: University of California Linguistic Minority Research Institute.
- Harklau, L. (1994). Jumping tracks: How language-minority students negotiate evaluations of ability. *Anthropology and Education Quarterly*, 25, 347–363. doi:10.1525/aeq.1994.25.3.04x0149s
- Heller, K. A., & Perleth, C. (2000). *Kognitiver Fähigkeitstest für 4. bis 12. Klassen, Revision: KFT 4–12+ R* [Test of Cognitive Abilities for Grades 4 to 12]. Weinheim: Beltz.
- Heller, V., & Morek, M. (2015). Academic discourse as situated practice: An introduction. *Linguistics and Education*, 31, 174–186. doi:10.1016/j.linged.2014.01.008
- Heppt, B., Stanat, P., Dragon, N., Berendes, K., & Weinert, S. (2014). Bildungssprachliche Anforderungen und Hörverstehen bei Kindern mit deutscher und nicht-deutscher Familiensprache [Academic language demands and listening comprehension in children with German and non-German home language]. *Zeitschrift für Pädagogische Psychologie*, 28, 139–149. doi:10.1024/1010-0652/a000130
- Hoff, E. (2013). Interpreting the early language trajectories of children from low-SES and language minority homes: Implications for closing achievement gaps. *Developmental Psychology*, 49, 4–14. doi:10.1037/a0027238
- Hu, L.-T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6, 1–55. doi:10.1080/10705519909540118
- Kieffer, M. J. (2010). Socioeconomic status, English proficiency, and late-emerging reading difficulties. *Educational Researcher*, 39, 484–486. doi:10.3102/0013189X10378400
- Köhne, J., Kronenwerth, S., Redder, A., Schuth, E., & Weinert, S. (2015) Bildungssprachlicher Wortschatz—Linguistische und psychologische Fundierung und Itementwicklung [Academic vocabulary—Linguistic and psychological foundations and development of items]. In A. Redder, J. Naumann, & R. Tracy (Eds.), *Forschungsinitiative Sprachdiagnostik und Sprachförderung—Ergebnisse* (pp. 67–92). Münster: Waxmann.
- Martiniello, M. (2008). Language and the performance of English-language learners in math word problems. *Harvard Educational Review*, 78, 333–368. doi:10.17763/haer.78.2.70783570r111t32
- Mason, A. (1995). By dint of: Student and lecturer perceptions of lecture comprehension strategies in first-term graduate study. In J. Flowerdew (Ed.), *Academic listening: Research perspectives* (pp. 199–218). Cambridge: Cambridge University Press.
- Meurers, D., Berendes, K., Vajjala, S., & Bryant, D. (2015). *Leseanforderungen in der Sekundarstufe: Ein Vergleich der linguistischen Komplexität von Schulbuchtexten* [Reading demands in secondary schools: A comparison of the linguistic complexity of textbooks]. Paper presented at the 3rd Tagung der Gesellschaft für Empirische Bildungsforschung, March. Retrieved from <http://purl.org/dm/handouts/gebf15.pdf>
- Morek, M., & Heller, V. (2012). Bildungssprache: Kommunikative, epistemische, soziale und interaktive Aspekte ihres Gebrauchs [Academic language: Communicative, epistemic, social, and interactive aspects of its utilization]. *Zeitschrift für angewandte Linguistik*, 2, 67–101. doi:10.1515/zfal-2012-0011
- Muthén, L. K., & Muthén, B. O. (1998–2012). *Mplus users guide* (7th ed.). Los Angeles: Author.
- Nagy, W., & Townsend, D. (2012). Words as tools: Learning academic vocabulary as language acquisition. *Reading Research Quarterly*, 47, 91–108. doi:10.1002/RRQ.011
- Nystrand, M. (2006). Research on the role of classroom discourse as it affects reading comprehension. *Research in the Teaching of English*, 40, 392–412.
- Organization for Economic Co-operation and Development. (2010). *PISA 2009 results: Overcoming social background—Equity in learning opportunities and outcomes* (Vol. 2). Paris: Author.
- Reese, L., Garnier, H., Gallimore, R., & Goldenberg, C. (2000). A longitudinal analysis of the antecedents of emergent Spanish literacy and middle-school English reading achievement of

- Spanish-speaking students. *American Educational Research Association Journal*, 37, 633–662. doi:10.3102/00028312037003633
- Richards, J. C. (1983). Listening comprehension: Approach, design, procedure. *TESOL Quarterly*, 17, 219–240. doi:10.2307/3586651
- Roessingh, H., Kover, P., & Watt, D. (2005). Developing cognitive academic language proficiency: The journey. *TESL Canada Journal*, 23, 1–27.
- Rost, M. (1994). *Introducing listening*. London: Penguin.
- Roth, H.-J., Neumann, U., & Gogolin, I. (2007). *Evaluation des Schulversuchs Bilinguale Grundschulen Hamburg: Abschlussbericht an die Behörde für Bildung und Sport Hamburg* [Evaluation of the pilot project Bilingual Elementary Schools in Hamburg]. Hamburg: Universität zu Hamburg.
- Saville-Troike, M. (1984). What really matters in second language learning for academic achievement? *TESOL Quarterly*, 18, 199–219. doi:10.2307/3586690
- Scarcella, R. (2003). *Academic English: A conceptual framework*. Santa Barbara, CA: University of California Linguistic Minority Research Institute.
- Scarcella, R., & Rumberger, R. W. (2000). Academic English key to long-term success in school. *University of California Linguistic Minority Research Institute Newsletter*, 9, 1–2.
- Schleppegrell, M. J. (2004). *The language of schooling: A functional linguistics perspective*. Mahwah, NJ: Erlbaum.
- Shaftel, J., Belton-Kocher, E., Glasnapp, D., & Poggio, J. (2006). The impact of language characteristics in mathematics test items on the performance of English language learners and students with disabilities. *Educational Assessment*, 11, 105–126. doi:10.1207/s15326977ea1102_2
- Shohamy, E., & Inbar, O. (1991). Validation of listening comprehension tests: The effect of text and question type. *Language Testing*, 8, 23–40. doi:10.1177/026553229100800103
- Snow, C. E. (2010). Academic language and the challenge of reading for learning about science. *Science*, 328, 450–452. doi:10.1126/science.1182597
- Snow, C. E., & Uccelli, P. (2009). The challenge of academic language. In D. R. Olson & N. Torrance (Eds.), *The Cambridge handbook of literacy* (pp. 112–133). Cambridge: Cambridge University Press.
- Tannen, D. (1982). The oral literate continuum of discourse. In D. Tannen (Ed.), *Spoken and written language* (pp. 1–6). Norwood, NJ: Ablex.
- Taylor, L., & Geranpayeh, A. (2011). Assessing listening for academic purposes: Defining and operationalising the test construct. *Journal of English for Academic Purposes*, 10, 89–101. doi:10.1016/j.jeap.2011.03.002
- Townsend, D., Filippini, A., Collins, P., & Biancarosa, G. (2012). Evidence for the importance of academic word knowledge for the academic achievement of diverse middle school students. *Elementary School Journal*, 113, 497–519. doi:10.1086/663301
- Uccelli, P., Barr, C. D., Dobbs, C. L., Galloway, E. P., Meneses, A., & Sánchez, E. (2015). Core academic language skills: An expanded operational construct and a novel instrument to chart school-relevant language proficiency in preadolescent and adolescent learners. *Applied Psycholinguistics*, 36, 1077–1109.
- Uccelli, P., Galloway, E. P., Barr, C. D., Meneses, A., & Dobbs, C. L. (2015). Beyond vocabulary: Exploring cross-disciplinary academic-language proficiency and its association with reading comprehension. *Reading Research Quarterly*, 50, 337–356. doi:10.1002/rrq.104
- Uessler, S., Runge, A., & Redder, A. (2013). “Bildungssprache” diagnostizieren: Entwicklung eines Instruments zur Erfassung von bildungssprachlichen Fähigkeiten bei Viert- und Fünftklässlern [Diagnosis of academic language: Development of an instrument for assessing academic language in four- and fifth-graders]. In A. Redder & S. Weinert (Eds.), *Sprachförderung und Sprachdiagnostik: Interdisziplinäre Perspektiven* (pp. 4–67). Münster: Waxmann.
- Vollmer, H. J., & Thürmann, E. (2010). Zur Sprachlichkeit des Fachlernens: Modellierung eines Referenzrahmens für Deutsch als Zweitsprache [Language in learning: Formulating a framework

- of reference for German as a second language]. In B. Ahrenholz (Ed.), *Fachunterricht und Deutsch als Zweitsprache* (pp. 107–132). Tübingen: Gunter Narr.
- Wald, A. (1943). Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Transactions of the American Mathematical Society*, 54, 426–482. doi:10.2307/1990256
- Weiß, R. H. (2006). *Grundintelligenztest Skala 2, CFT 20—Revision* [Culture Fair Intelligence Test—Scale 2]. Göttingen: Hogrefe.
- Wilson, M. (2005). *Constructing measures: An item response modeling approach*. Mahwah, NJ: Erlbaum.
- Wolf, M. K., Crosson, A. C., & Resnick, L. B. (2005). Classroom talk for rigorous reading comprehension instruction. *Reading Psychology*, 26, 27–53. doi:10.1080/02702710490897518
- Wu, M. L., Adams, R. J., Wilson, M. R., & Haldane, S. A. (2007). *ACER ConQuest Version 2: Generalised item response modelling software*. Camberwell, Australia: ACER Press.